

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Correlation

Lecture 48 Section 13.7

Robb T. Koether

Hampden-Sydney College

Mon, Apr 21, 2008

Outline

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- 1 Introduction
- 2 The Correlation Coefficient
- 3 Calculating the Correlation Coefficient
- 4 Calculating r on the IT-83
- 5 Hidden Variables
- 6 Summary

Introduction

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- So far our only way to judge the strength and direction of a linear relationship has been to look at a scatterplot.
- Now we will quantify those concepts in a quantify called the correlation coefficient.
- However, there are some dangers present.
 - 1 “Correlation does not imply causation.”
 - 2 There may be hidden variables at work.

The Correlation Coefficient

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- The **correlation coefficient** r is a number between -1 and $+1$.
- It measures the direction and strength of the linear relationship.
- If $r > 0$, then the relationship is positive. If $r < 0$, then the relationship is negative.
- The closer r is to $+1$ or -1 , the stronger the relationship.
- The closer r is to 0 , the weaker the relationship.

Strong Positive Linear Association

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

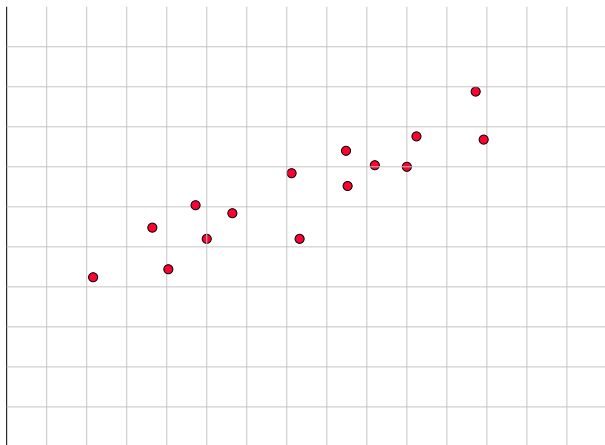
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

In this display, r is close to $+1$



Strong Negative Linear Association

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

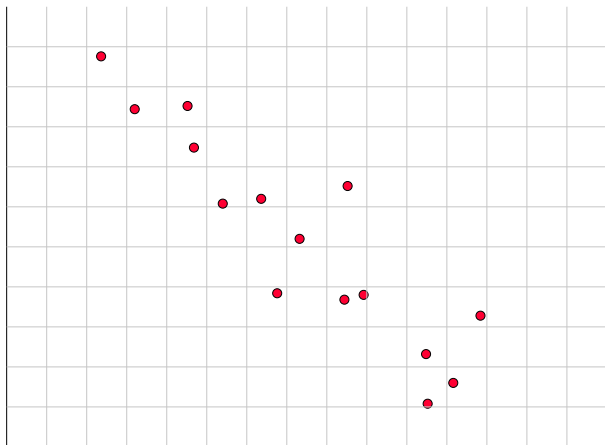
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

In this display, r is close to -1



Almost No Linear Association

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

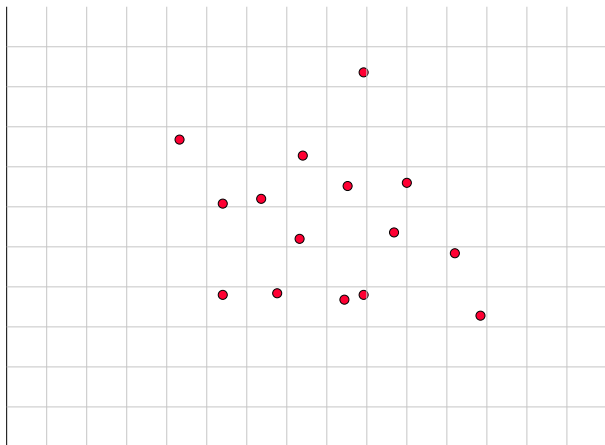
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

In this display, r is close to 0



Interpretation of r

Correlation

Robb T.
Koether

Introduction

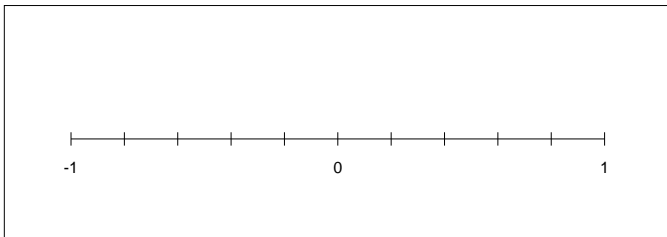
The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary



Interpretation of r

Correlation

Robb T.
Koether

Introduction

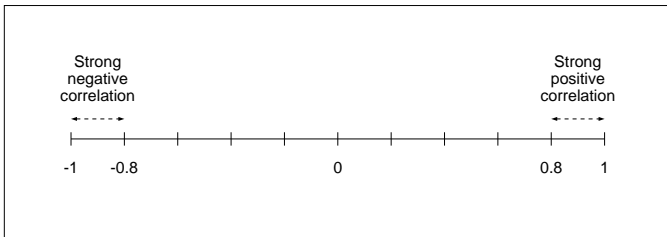
The Correlation Coefficient

Calculating the Correlation Coefficient

Calculating r on the IT-83

Hidden Variables

Summary



Interpretation of r

Correlation

Robb T.
Koether

Introduction

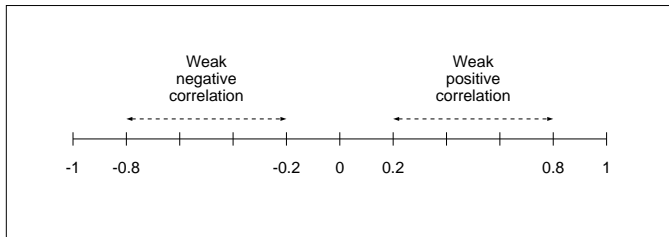
The Correlation Coefficient

Calculating the Correlation Coefficient

Calculating r on the IT-83

Hidden Variables

Summary



Interpretation of r

Correlation

Robb T.
Koether

Introduction

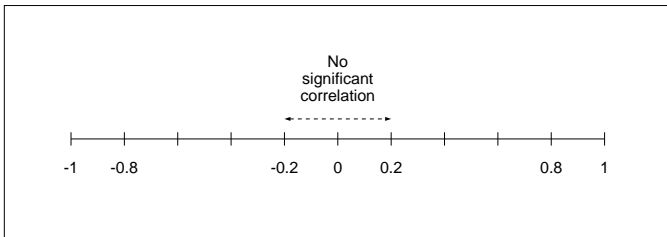
The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary



Correlation vs. Cause and Effect

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- If the value of r is close to $+1$ or -1 , that indicates that x is a good **predictor** of y .
- It does not indicate that x **causes** y .
- The correlation coefficient alone cannot be used to determine cause and effect.

Calculating the Correlation Coefficient

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- There are many formulas for r .
- The basic formula is

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

which can be expanded to

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}.$$

Calculating the Correlation Coefficient

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- If we define

$$SSX = \sum (x - \bar{x})^2,$$

$$SSY = \sum (y - \bar{y})^2,$$

$$SSXY = \sum (x - \bar{x})(y - \bar{y}),$$

then the formula becomes

$$r = \frac{SSXY}{\sqrt{SSX \cdot SSY}}.$$

Example

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider again the data

x	y
1	8
3	12
4	9
5	14
8	16
9	20
11	17
15	24

Example

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- We found earlier that

$$SSX = 150$$

$$SSY = 206$$

$$SSXY = 165$$

Example

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Then compute r :

$$r = \frac{165}{\sqrt{(150)(206)}} = 0.9387.$$

Calculating r on the IT-83

Correlation

Robb T.
Koether

Introduction

The Correlation Coefficient

Calculating the Correlation Coefficient

Calculating r on the IT-83

Hidden Variables

Summary

- To calculate r on the TI-83,
 - Be sure that `Diagnostic` is turned on.
 - Press `CATALOG` and select `DiagnosticsOn`.
 - Then, follow the procedure that produces the regression line.
 - In the same window, the TI-83 reports the values of r^2 and r .

Calculating r on the IT-83

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Use the TI-83 to calculate r in the preceding example.
- Find r for English SOL Passing Rate vs. Graduation Rate.
- Find r for the Free Lunch vs. Graduation Rate.
- Find r for the Teachers' Salary vs. Graduation Rate.
- Find r for the S/T Ratio vs. Graduation Rate.

Calculating r on the IT-83

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- For the Free Lunch vs. Graduation Rate, we get $r = -0.8544$.
- For English SOL Passing Rate vs. Graduation Rate, we get $r = 0.7500$.
- For the Teachers' Salary vs. Graduation Rate, we get $r = 0.0817$.
- For the S/T Ratio vs. Graduation Rate, we get $r = 0.0022$.

Calculating r on the IT-83

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- For the Free Lunch vs. Graduation Rate, we get $r = -0.8544$.
- For English SOL Passing Rate vs. Graduation Rate, we get $r = 0.7500$.
- For the Teachers' Salary vs. Graduation Rate, we get $r = 0.0817$.
- For the S/T Ratio vs. Graduation Rate, we get $r = 0.0022$.

Calculating r on the IT-83

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- For the Free Lunch vs. Graduation Rate, we get $r = -0.8544$.
- For English SOL Passing Rate vs. Graduation Rate, we get $r = 0.7500$.
- For the Teachers' Salary vs. Graduation Rate, we get $r = 0.0817$.
- For the S/T Ratio vs. Graduation Rate, we get $r = 0.0022$.

Calculating r on the IT-83

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- For the Free Lunch vs. Graduation Rate, we get $r = -0.8544$.
- For English SOL Passing Rate vs. Graduation Rate, we get $r = 0.7500$.
- For the Teachers' Salary vs. Graduation Rate, we get $r = 0.0817$.
- For the S/T Ratio vs. Graduation Rate, we get $r = 0.0022$.

Another Formula for r

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the TI-83

Hidden
Variables

Summary

- Another formula for r is

$$r = \frac{1}{(n-1)} \sum \left(\frac{x - \bar{x}}{s_X} \right) \left(\frac{y - \bar{y}}{s_Y} \right).$$

- This formula is based on the z-scores of x and y .

Another Formula for r

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- For the following data set, use the z-scores of x and y to find r .

x	y
1	8
3	12
4	9
5	14
8	16
9	20
11	17
15	24

The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Suppose we investigate the relationship in children (3rd - 5th grade) between number of hours spent watching TV and their reading level.
- One would expect a negative relationship.

The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

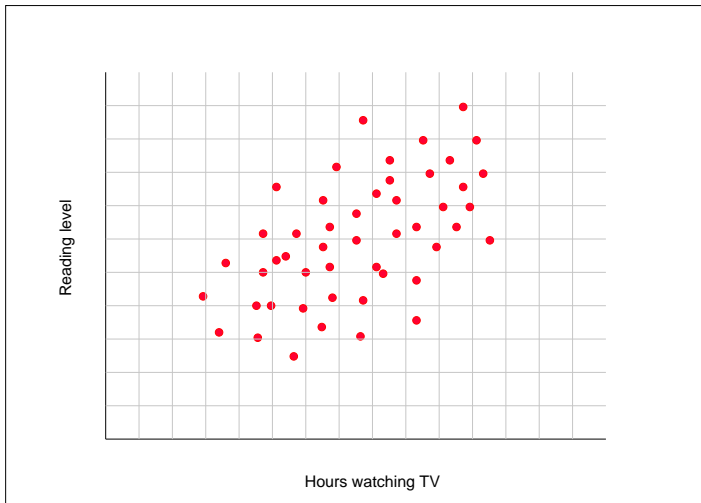
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Weak Positive Correlation



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- There appears to be a weak positive relationship between the two variables!
- What an amazing discovery! Tell the kids to watch more TV!
- But now consider a possible third variable: grade level.
- Divide the sample into three groups according to grade level (3rd, 4th, and 5th).

The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

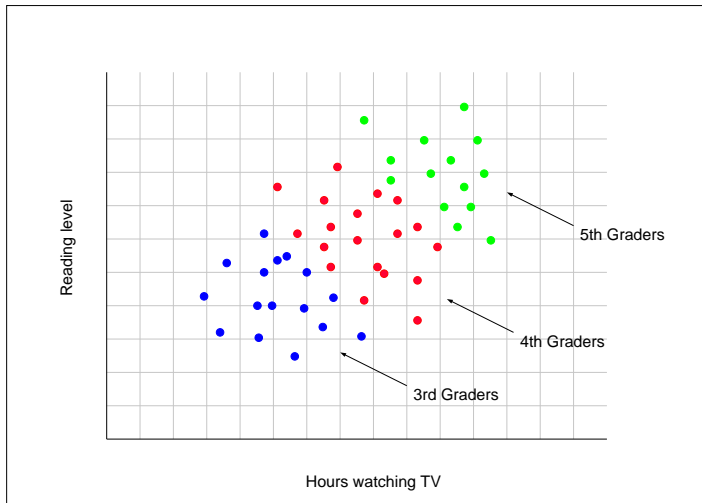
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Weak Positive Correlation?



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

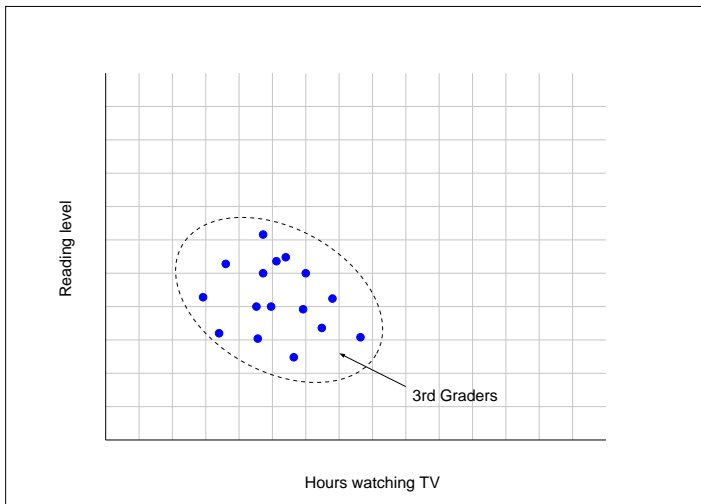
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Weak Negative Correlation Among 3rd Graders



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

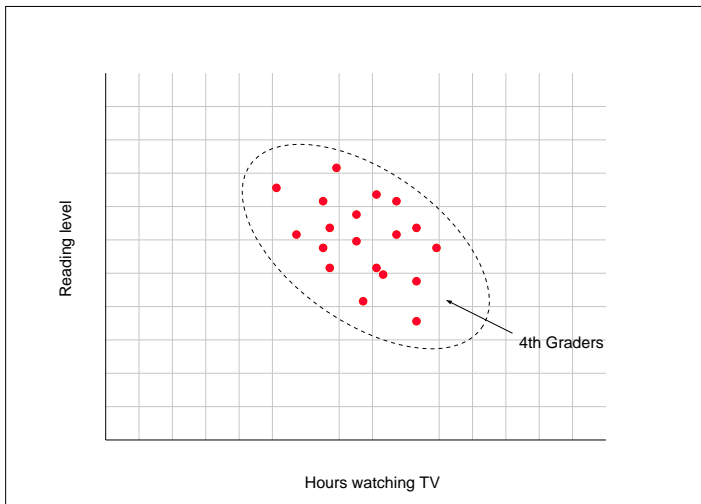
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Weak Negative Correlation Among 4th Graders



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The Correlation Coefficient

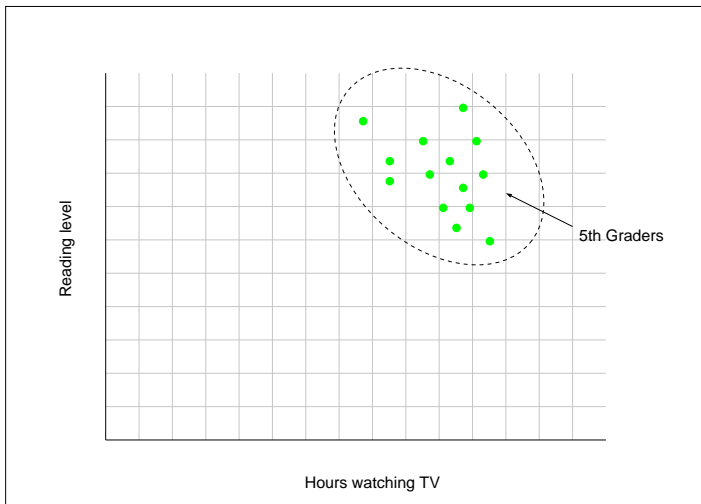
Calculating the Correlation Coefficient

Calculating r on the IT-83

Hidden Variables

Summary

Weak Negative Correlation Among 5th Graders



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the famous example of the strong positive correlation between one's score on the Math SAT and one's shoe size.
- What could be a possible hidden variable?

The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

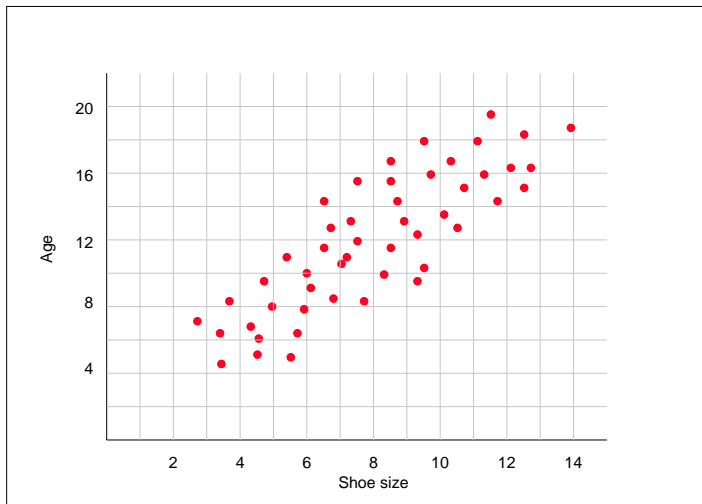
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Shoe size is a predictor of age



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

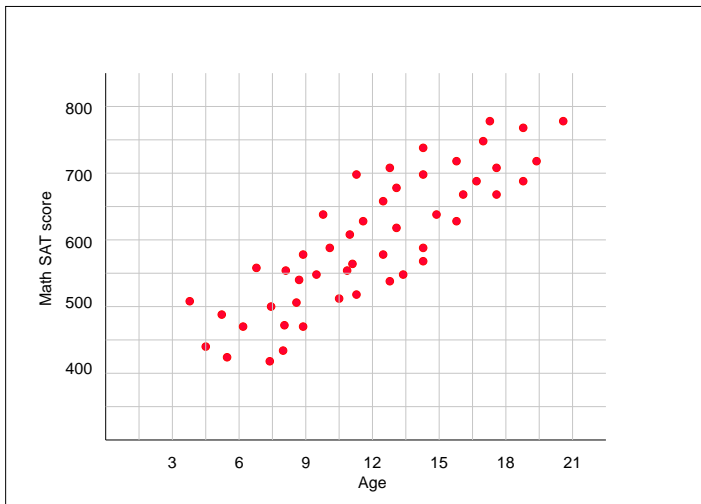
Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

**Hidden
Variables**

Summary

Age is a predictor of Math SAT score



The Danger of Hidden Variables

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Article

Another Mostly Useless Correlation

Another Formula for r

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Free-lunch participation vs. graduation rate data, SSR = 1896.7, SST = 2598.2.
- So we get

The Coefficient of Determination

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- r^2 is called the coefficient of determination.
- It is interpreted as telling us how much of the variation in y is determined by the variation in x .
- So, 73% of the variation in graduation rates is determined by the variation in participation in the free-lunch program.

The Coefficient of Determination

Correlation

Robb T.
Koether

Introduction

The Correlation Coefficient

Calculating the Correlation Coefficient

Calculating r on the IT-83

Hidden Variables

Summary

- What percentage of the variation in graduation rate is determined by the variation in S/T ratio?
- What percentage of variation in graduation rate is determined by the variation in teachers' average salary?

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- How does r indicate the direction of the relationship?
- Consider the numerator of the formula.

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

How Does r Work?

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

- Consider the lunch vs. graduation data:

Scatter Plot

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Scatter Plot

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Scatter Plot

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Scatter Plot

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Summary

Correlation

Robb T.
Koether

Introduction

The
Correlation
Coefficient

Calculating
the
Correlation
Coefficient

Calculating r
on the IT-83

Hidden
Variables

Summary

Summary

